Docket No. 2799/64201

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

(U.S. DESIGNATED OFFICE)

APPLICATION FOR

UNITED STATES LETTERS PATENT

BASED UPON:
INTERNATIONAL APPL. NO. PCT/NL99/00518
UNDER PATENT COOPERATION TREATY

INVENTOR: Otte, Arie Pieter

TITLE: Method for detecting a DNA sequence, a DNA sequence, a method of making a DNA construct and the use thereof

ATTORNEY: Robert D. Katz
Reg. No. 30,141
COOPER & DUNHAM LLP
1185 Avenue of the Americas
New York, New York 10036
(212) 278-0400
Attorney for Applicant

Method of detecting a DNA sequence, a DNA sequence, a method of making a DNA construct and the use thereof

The present invention relates to a method of detecting, and optionally selecting, a DNA sequence.

It is not easy to detect a specific DNA sequence of which the nucleotide sequence is not known. Despite the fact that genetic manipulation has been employed for decades, predictably bringing to expression a gene in a genetically modified plant, animal or other eukaryotic organism is a problem. Although many microbiological methods of production merely aim at the highest possible expression, in plants or animals the exact level of a gen's expression is for many applications of great importance. Too much expression as well as too little expression may lead to the desired result not being achieved. Also, experience has shown that after sexual reproduction the ability for expression in a subsequent generation is often lost again. It is also difficult to control the moment in time and the location of expression in the organism (tissue specificity).

It is the object of the invention to provide a method of the kind mentioned in the preamble, which makes it possible to select and, if desired, isolate a DNA sequence, whereby the above-mentioned problems can be avoided.

To this end the method according to the preamble is characterized in that the DNA sequence to be detected possesses a stable expression-enhancing quality, which method comprises the steps of

1) the cloning in a vector of DNA fragments having a size of <5000 base pairs between i) a DNA sequence involved in the induction of gene transcription-repressing chromatin, and ii) a reporter gene comprising a promotor, resulting in a variety of a fragment-comprising vectors, wherein the distance between the DNA sequence involved in the induction of the transcription of

gene-repressing chromatin and the reporter gene is fewer than 5000 base pairs;

2) introducing the vectors into host cells, in which host cells the promotor may be active but induction of the transcription of gene-repressing chromatin in the vectors results in the repression of the transcription of the reporter gene; and

3) subjecting the host cells to a selection in order to identify a host cell exhibiting reporter gene-activity.

This provides a reliable method of detecting DNA sequences having a stable expression-enhancing quality. If desired, this sequence may be isolated and inserted before another gene. As the DNA in step 1, for example, a restriction enzyme-cleaved DNA from a eukaryotic organism, in particular a plant or a vertebrate, is used wherein the size of the DNA fragments is below 5000 base pairs.

Clearly, when the occasion arises it will be possible to readily distinguish between on the one hand an expression-enhancing sequence ("enhancer"), which in extreme cases would be able to neutralize the transcription-repressing effect of chromatin, and on the other hand the stable expression-enhancing DNA fragment. In the first case the reporter gene in an organism is transformed with a vector comprising the promotor together with the reporter gene but without the transcription-repressing sequence is expressed at a higher level than in an organism transformed with a vector comprising a stable expression-enhancing DNA fragment together with the reporter gene and likewise, without the transcription-repressing sequence.

According to a first preferred embodiment, the selection in step 3) occurs by using a reporter gene which provides resistance to a growth inhibitor and the host cells are cultivated in the presence of the growth inhibitor.

This inhibits the growth of host cells which, without an active resistance gene, are not resistant to the growth inhibitor, and allows the selection of those host

cells which possess a stable expression-enhancing DNA sequence.

Preferably, the growth inhibitor is present in a concentration sufficiently high to kill host cells in which the gene providing resistance to the growth inhibitor is not active.

This ensures to a large extent that growing organisms will comprise a vector with the desired DNA sequence.

Very conveniently an antibiotic is used as the growth inhibitor and the reporter gene is a gene providing resistance to the antibiotic.

A great assortment of genes providing resistance to antibiotics is available in the field, making it simple to choose a gene suitable for the host cell. A gene is then chosen which provides resistance to a growth inhibitor to which the host cell is not already resistant of itself.

In accordance with a second embodiment the reporter gene codes for Green Fluorescent Protein.

By means of fluorescence measurement it is then possible to detect and isolate host cells with the desired DNA-comprising vector.

According to a preferred embodiment, fluorescent host cells are separated from non-fluorescent host cells by means of a Fluorescence-Activated Cell Sorter (FACS).

According to a third embodiment the reporter gene is luciferase. With the aid of luciferase it is possible to perform (semi)-quantitative measurement of the expression.

In step 1) it is preferred that the fragments have a size of substantially between 2000 - 3000 base pairs.

Fragments of such a size allow a more precise localization of the sequence to be detected without the number of host cells to be screened in step 3) becoming so large that this is going to form an unnecessary work load.

Conveniently, the DNA sequence involved with the transcription induction of gene-repressing chromatin is a DNA sequence that is recognized by a heterochromatin-binding protein comprising HP1 (heterochromatin-binding protein 1), which HP1-comprising complex is expressed in the

host cell. According to an alternative method, the DNA sequence is recognized by a complex comprising a Polycomb-group (Pc-G) protein, and the Polycomb-group protein-comprising complex is expressed in the host cell. According to yet another embodiment, the DNA sequence is recognized by a complex possessing a histone deacetylase activity, and the histone deacetylase activity-possessing complex is expressed in the host cell. Finally, according to a further embodiment, the DNA sequence involved in the induction of the transcription of gene-repressing chromatin is a DNA sequence recognized by a protein complex comprising MeCP2 (methyl-CpG-binding protein 2), and the MeCP2-comprising complex is expressed in the host cell.

In this manner four suitable complexes recognizing DNA sequences are provided, while it should be noted that in the event of the complex not being expressed in the host cell, this will not result in false positives and will merely limit the efficiency with which the wanted DNA sequences are detected.

Conveniently, the protein complex comprises a fusion protein, such as a protein complex wherein the first part is a part binding the DNA-binding site of LexA-DNA or GAL4-DNA.

Suitable DNA binding sites of this kind are known in the art and are obtained from bacteria or yeast.

The organism in step 1) is preferably chosen from the group comprising a plant and a vertebrate such as, more particularly, a mammal.

For these organisms applies that, partly due to the large amount of chromosomal DNA, it is practically impossible without the method of the present invention to find the DNA sequence to be detected, since indeed its base sequence is unknown.

According to a further preferred embodiment, the vector is an episomally replicating vector, such as suitably a vector comprising a replication origin from the Epstein-Barr virus (EBV), OriP, and a nuclear antigen (EBNA1).

Such vectors are easy to handle, can be genetically manipulated and are vectors which form a chromatin structure in which the expression is repressed.

The invention further relates to a DNA sequence selected from i) a DNA sequence isolated from a plant or vertebrate, or derivatives thereof, and ii) a synthetic DNA sequence or one constructed by means of genetic engineering, which DNA sequence is a repression-inhibiting sequence which, by the method according to the present invention can be detected, selected and optionally cloned.

More specifically, the invention further relates to a DNA sequence selected from i) a DNA sequence isolated from a plant or vertebrate, or derivatives thereof, and ii) a synthetic DNA sequence or one constructed by means of genetic engineering, which DNA sequence is detected, selected and optionally cloned by the method according to the present invention.

The DNA sequences according to the invention differ from the known DNA sequences in that they are not an enhancer or silencer.

Synthetic DNA sequences may be prepared in accordance with techniques generally known in the art. In particular, it is possible to prepare large numbers of different DNA sequences, and such sequences are commercially available (for example from: Pharmacia Biotech, Uppsala, Sweden). However, such synthetic DNA sequences have to be suitable for cloning in a plasmid. This is generally known in the art and is done, for example, with linkers comprising a restriction cleavage site.

Clearly, the present invention also relates to a method of making a DNA construct comprising a gene that is to be expressed stably, wherein a stable expression-enhancing DNA sequence, selected with the aid of the method according to the invention is inserted at less than 2000 bp from the gene.

This is a more stable and predictable manner of expressing a gene.

Preferably the stable expression-enhancing DNA sequence will be inserted both upstream and downstream from the gene.

It is believed that this further increases the likelihood of a stable gene expression.

Finally, the invention relates to a use of the DNA construct according to the invention, wherein the DNA construct is a vector, for the transformation of an organism which suitably is an organism as defined above.

The present invention will now be further elucidated with reference to the following exemplary embodiments.

Example I

To illustrate the principle of the workings of the method according to the invention, scs is used, which is a DNA fragment from *Drosophila melanogaster* which is known to be a boundary element. As can be seen from the example below, scs can be used for blocking the following repressors: HP1, Polycomb-group proteins and MeCP2. In the same manner, DNA fragments from phage lambda have been tested as negative control. Scs (special chromatin structure) was originally isolated as a DNA sequence flanking the heat shock locus (*hsp*70) in *Drosophila* (Kellum, R. and P. Schedl. 1991. Cell 64: 941-950). They have found that when scs is placed around a reporter gene and is reintroduced in *Drosophila*, the expression of a reporter gene is less variable. They neither reported nor suggested that scs may be used to prevent repression by other repressors, in particular the above-mentioned repressors. Also, Kellum et al. neither reported nor suggested that scs might be used in systems other than *Drosophila* for rendering transgene expression less variable.

For testing the repression-eliminating property of a DNA sequence, two types of vectors are constructed.

The first type of vector comprises in 5'-3' sequence: four LexA binding sites, the scs sequence to be tested, the human heat shock factor-inducible promotor, and the luciferase gene as reporter gene. As a control a

similar vector is made which instead of the known scs sequence comprises a random fragment (from phage lambda) of a comparable length (both described in point 1 below).

To accomplish repression of the reporter gene in the transformed cell, the second type of vector comprises a gene coding for a fusion protein of LexA and the above-mentioned repressors. A vector of this second type comprises the gene coding for LexA only, or a vector comprises the gene coding for LexA-HP1, etc. (described in point 2 below).

1 A vector coding for EBNA-1 (a nuclear antigen) is the hygromycin resistance gene comprising pREP4 vector (Invitrogen Corporation, Carlsbad, USA). De EBNA-1 sequence is present to ensure that the vector does not (stably) integrate in the genome, but replicates episomally. The promotor (Prsv) of this vector has been removed by digestion with the restriction enzyme SalI and replaced by a synthesized sequence having four binding sites for LexA from *E. coli*. This sequence is from 5'-3': **GTCGACTGCTGTATATAAAACCAGTGGTTATATGTACAGTACTT** GTACTGTACATATAACCACTGGTTTTATATACAG-**CAAGCTTGGATCCGTCGAC**. The 5' side of this sequence comprises a SalI site, the 3' side a HindIII-BamHI-SalI site (all shown in bold type). Downstream from the LexA binding sites in the HindIII and BamHI sites, the human heat shock factor-inducible promotor (0.29 kbp HindIII/NcoI fragment) and the luciferase reporter gene inclusive of SV40 polyadenylation signal (1.9 kbp NcoI/BamHI fragment) are cloned in a three-way ligation. The human heat shock factor-inducible promotor (hsp70; accession numbers M59828 and M34267; nucleotides 52 to 244) can be obtained by means of PCR amplification on human genomic DNA (Cat. No. 6550-1; Clontech, Palo Alto, USA). As PCR primers, forward primer 5'-3': **AAGCTTGGGAGTCGAAACTTCTGGAATATTCCCGAACTTTCAGCCGACGA**-CTTATAAAACGCCAGGGGCAAGC may be considered; and as reverse primer 5'-3': **CCATGGTTTAGCTTCCTTAGCTCCTGAA**-

AATCTCGCCAAGCTCCCGGGGTCCGCGAGAAGAGCTCGGTCCTTCCGG. The forward primer comprises a HindIII site, the reverse primer comprises a NcoI site (given in bold print). The luciferase reporter gene inclusive of SV40 polyadenylation signals were obtained through NcoI/BamHI digestion of the pGL3 control vector (Cat. no E1741; Promega, Madison, USA). In the thus obtained vector, in the HindIII site between the LexA binding sites and the heat shock promotor, either a 2.1 kbp HindIII fragment of phage lambda is cloned (Pharmacia Biotech, Uppsala, Sweden), or a 1.7 kbp scs HindIII fragment. The 1.7 kbp scs DNA fragment is isolated from genomic *Drosophila* DNA (Cat.#6940-1, Clontech, Palo Alto, USA) with the aid of PCR primers (Forward primer 5'-3': GATCAAGCTTATGATCTGCGTATGATACCAAATTTCTG; Reverse primer 5'-3': GACAAGCTTACATTGCTGGGCGAGCTGCGCCAATCG). At the ends of these primers HindIII restriction enzyme sites were located. The vector with the Lambda fragment (control) is indicated as reporter construct a, the vector with the scs fragment as reporter construct b. Restriction enzyme digestions, PCR amplifications and clonings are performed by standard procedures as described in Sambrook et al., Molecular Cloning; a laboratory manual, second edition.

2 The DNA-binding domain of the LexA protein (aa 1-202) (Cat.#6183-1, Clontech, Palo Alto, USA) is cloned in the HindIII site of the neomycin resistance gene-comprising pREP9 (Invitrogen Corporation, Carlsbad, USA) vector. Downstream and in frame with the LexA gene, one gene coding for a repressor is cloned per vector. The repressors used are: the 1674 bp-long coding part of the humane Polycomb-group gene HPC2 (accession number Genbank: AAB80718), the 1131 bp-long coding part of the humane Polycomb-group gene RING1 (accession number Genbank: Z14000), the 4098 bp-long coding part of

the *Drosophila* Polycomb-group gene Su(z)2 (accession number Genbank: CAA41965), the 558 bp coding part of *M31 (mHP1)* (accession number Genbank: P23197), or the 1478 bp coding part of *MeCP2* (accession number Genbank: A41907). These constructs code for LexA-HPC2, LexA-RING1, LexA-Su(z)2, LexA-mHP1 and LexA-MeCP2 fusion proteins, or **LexA repressors**. These bind to the LexA binding sites (see point 1).

3 The reporter vectors a and b and the LexA repressor-coding vectors are expressed in humane U-2 OS (osteosarcoma) cells obtained from the ATTC (accession number HTB-96). Transfection of the cells with the DNA constructs is performed using the calcium phosphate method in accordance with the instructions of the manufacturer of the transfection kit (Cat. No. 18306-019, Gibco BRL, Gaithersburg, USA). The osteosarcoma cells grow in the presence of 100 µg/ml neomycin (G418: Cat. No. 1464981; Boehringer/Roche, Switzerland) and 50 µg/ml hygromicin B (Cat. No 843555; Boehringer/Roche, Switzerland). Three days after transfection a heat shock is given (43°C for 1 hour, followed by a 6-hour recovery period at 37°C). This treatment activates the luciferase gene and causes the production of the luciferase reporter protein. The enzymatic activity of this luciferase protein is a measure of the transcription induction that has been induced. Cells are purified and the luciferase enzyme activity is measured, all in compliance with the manufacturer's instructions for the standard luciferase reporter gene assay kit (Cat. No. 1814036; Boehringer/Roche, Switzerland).

Result

4 In cells in which the reporter construct a (with the Lambda fragment) is expressed, but no LexA

repressors, the luciferase gene is expressed after heat shock. This is the 100% value.

5 In cells in which the reporter construct b (with the scs fragment) is expressed, but no LexA repressors, the luciferase gene is expressed after heat shock up to a value of 100%. Since this value does not exceed the 100% it shows, as explained earlier, that it is not an expression-increasing sequence.

6 In cells in which the reporter construct a (with the Lambda fragment) is expressed, and also LexA repressors are expressed, the expression of the luciferase gene after heat shock is repressed to an average of 20%.

7 In cells in which the reporter construct b (with scs fragment) is expressed, and at the same time LexA repressors, the expression of the luciferase gene after heat shock reaches a value of 100%. This shows that the induction of the repressor activity can be repressed with scs.

Example II

Instead of luciferase as reporter gene, it is according to the present invention also possible to use another reporter gene. It is also possible to use other promotors.

8 In the reporter constructs a and b the luciferase reporter gene has been replaced by the Zeocin resistance gene. The heat shock promotor has been replaced by the constitutive SV40 promotor (pSV40/ZEO; Cat. No. V502-20; Invitrogen, Carlsbad, USA). After transfection the U-2 OS cells grow in 250 µg/ml Zeocin (Cat. No. R250-01: Invitrogen, Carlsbad, USA) and 100 µg/ml neomycin (G418: Cat. No. 1464981; Boehringer/Roche, Switzerland).

9 Cells that have been transfected with the selection construct comprising a 2.1 kbp Lambda fragment and also with a construct that expresses a LexA repressor, die after 20-30 days. This shows that

the Lambda fragment is not able to overcome the repression of the gene with which antibiotics resistance is achieved.

10 Cells that are transfected with the selection construct comprising the scs fragment and also with a construct that expresses a LexA repressor, do not die but continue to grow. This also shows that with the boundary element scs the repression can be overcome and that the method according to the present invention can be employed using a variety of promotors and reporter genes.

Example III

The sequences found and selected by the method according to the invention can be used to combat repression in an organism other than that from which the sequence is derived.

11 Two new constructs, c and d, are made, so-called T-DNA constructs, which are suitable for the transformation of plants.

12 Construct c comprises a cassette with the NPTII (neomycin phosphotransferase II) gene for resistance selection with kanamycin and the GUS (β-glucuronidase) reporter gene. The NPTII gene is regulated by the constitutive nos promotor and the GUS reporter gene by the constitutive CaMV 35S promotor (Mlynarova, L. et al., 1995. The Plant Cell 7: 599-609).

13 Construct d is construct c in which an scs fragment is cloned immediately upstream from the GUS-CaMV/nos-NPTII cassette and an scs fragment immediately downstream from the cassette.

14 Agrobacterium tumefaciens is transformed with construct c or d. *Arabidopsis* plants are submerged in a suspension (culture) of *Agrobacterium tumefaciens* with construct c and in a suspension of *Agrobacterium tumefaciens* with construct d (Clough et al., 1998. The Plant J. 16: 735-743).

15 40 individual *Arabidopsis* plants with construct c or d are raised and the seeds of the plants collected. The seeds are sown onto a medium containing kanamycin (Cat. No. 106801; Boehringer/Roche, Swiss) and GUS reporter activity is measured in the leaves of the developed plants.

16 The GUS activity in plants with construct c is very variable (7 high; 6 intermediate; 11 low; 16 zero); the GUS activity in plants with construct d is systematically higher and the variability is reduced (26 high; 4 intermediate; 5 low; 5 zero).

17 This shows that a gene can be expressed more stably with a boundary element, even if this boundary element does not originate from the same organism.